Akshu Chahar
National Institute of Technology, Raipur
June 03, 2021 - July 02, 2021

Project 1 - Twitter Data Extraction

The project aims at gathering data or tweets from Mumbai or Pune and analyse it to know about the food trends regarding home cooked food and tiffin services. This helps in knowing the future trend in home food.

What I did :

- Searched for popular hashtags across Twitter.

["mumbai tiffins", "pune tiffins", "#Tiffinservice", "#tiffinforcorporate", "#onlinetiffin", "#tiffinserviceindia", "#onlinetiffinservice", "#tiffin", "#TIFFIN"]

- Searched for keywords/hashtags used by various popular tiffin services across Mumbai and Pune. The popular services were:

["Mumbai Dabbawala", "Spice Box", "Yummy Tiffins", "Happy Grub"].

Learnings:

- The usage of Twitter Developer's Account and Tweepy.
- How to extract tweets from Twitter and perform EDA on it.
- Searching for keywords.

Inference:

Most of the famous tiffin services are not much active on twitter. Hence, the data for analysis and prediction of food trends needs to be acquired from consumers' tweets.

Akshu Chahar
National Institute of Technology, Raipur
June 03, 2021 - July 02, 2021

Project 2 - Meal Prices Prediction

This project aims at building a model similar to Mumbai Housing Prices prediction model which will be able to predict the Meal Prices in different areas of Mumbai. This was done by using **Kaggle Dataset** Mumbai1.csv.

What I did :

- Data Cleaning:

The major problem in the dataset was the representation of the same location in different ways. **For example,** Kharghar was represented as Kharghar as well as with some more exact location by providing the Sector (such as Sector 13 Kharghar).

Then, focusing on each of the locations, the second problem was the outliers. These were taken care of by focusing on **only 1, 2 and 3 bedroom** houses.

- Building Model and Evaluation (Location-wise):

By using **RandomForestRegressor**, the housing prices were predicted for each of the selected locations (**“Kharghar”, “Mira Road”, “Borivali West”, “Malad West”, “Thane West”, “Andheri East”, “Andheri West”**). It was found that the model was working nicely with a **R2 score of 75-85%**. Although “Andheri East” and “Andheri West” were neglected for further processing as there was not enough data for them.

- Final Model:

Kharghar, Mira Road, Borivali West and Malad West were used as training dataset with meal prices 80, 80, 80 and 70 respectively. **Thane West was used as a test dataset**.

Learnings:

- Data Cleaning techniques.
- Model Evaluation by looking at the predicted values.

Inference:

The model worked well with House price predictions because of the continuous nature of House prices but failed while predicting the meal prices because of their categorical nature. This made it a Classification problem instead of a Regression problem.